

Formulation of the augmented plane-wave and muffin-tin orbital method

Takao Kotani

Department of applied mathematics and physics, Tottori university, Tottori 680-8552, Japan

Hiori Kino

National Institute for materials science, Sengen 1-2-1, Tsukuba, Ibaraki 305-0047, Japan.

Hisazumu Akai

Institute of solid state Physics, Tokyo university, Kashiwa 277-8581, Japan

(Dated: October 28, 2014)

A mixed basis all-electron full-potential method, which uses two kinds of augmented waves, the augmented plane waves and the muffin-tin orbitals simultaneously, in addition to the local orbitals, was proposed by Kotani and van Schilfgaarde in Phys. Rev. B81, 125117(2010). We named it the PMT method. In this paper, this mixed basis method is reformulated on the basis of a new formalism named as the 3-component formalism, which is a mathematically transparent version of the additive augmentation originally due to Soler and Williams in Phys. Rev. B47, 6784(1993). Atomic forces are easily derived systematically. We discuss some problems in the mixed basis method and ways to manage them. In addition, we compare the method with the PAW method on the same footing. This PMT method is the basis for our new development of the quasiparticle self-consistent *GW* method in J.Phys.Soc.Jpn 83, 094711(2014).

PACS numbers: 71.15.Ap, 71.15.-m, 31.15.-p

I. INTRODUCTION

In the first-principles electronic-structure calculations based on the density functional theory in the LDA/GGA (local density approximation/generalized gradient approximation), a key element is the one-body problem solver, which should have efficiency, accuracy and robustness. As such solvers, the linearized augmented plane wave (LAPW) method and the linearized muffin-tin orbital (LMTO) method were proposed by Andersen in 1975 [1], followed by many improvements and extensions [2–7]. Today LAPW and LMTO has developed to be full-potential methods, which we treat in this paper. In these methods, wavefunctions are represented by superpositions of augmented waves. The LAPW uses the augmented plane waves (APWs) made of plane waves (PWs) as envelope functions; the LMTO uses the muffin-tin orbitals (MTOs) made of the atom-centered Hankel functions. Corresponding to these envelope functions, the APWs fit to the extended nature of wavefunctions; in contrast, the MTOs to the localized nature of them. However, wavefunctions in real materials should have both the natures.

This fact is reflected as shortcomings in these methods. In the case of the LAPW, it requires so many bases in order to represent sharp structures of wavefunctions just outside of muffin-tins. For example, *3d* orbitals of transition metals are the typical cases. Most of all the PWs used in the LAPW method are consumed only to reproduce the sharp structures. On the other hand, the LMTO is problematic in representing the extended nature of wavefunctions. For example, we sometimes need to put empty spheres between atoms. In addition, it is not simple to enlarge basis set systematically in order to

check numerical convergence.

To overcome these shortcomings, Kotani and van Schilfgaarde introduced a new linearized method named as the APW and MTO method (the PMT method) [8], which is an all-electron (AE) full-potential mixed basis method using APWs and MTOs simultaneously. Because these two kinds of basis are complementary, we can overcome these shortcomings. Within our knowledge, no other mixed basis methods have used different kinds of augmented waves simultaneously in the full-potential methods.

A minimum description on the formalism of the PMT method is given in Ref.[8], which is based on Ref.[5] for a LMTO method by Methfessel et al. However, the formalism was not very transparent, mainly because it was not derived from the explicit total energy minimization. This makes theoretical treatment of the PMT method somehow complicated. For example, it resulted in a complicated logic to derive atomic forces in Refs.[5, 9]. It was not easy to compare the PMT method with the projector augmented wave (PAW) methods [6, 10] on the same footing. Thus we should give a simple and clear formalism to the PMT method for its further developments rather than that in Refs.[5, 8].

In this paper, we introduce a formalism, named as the 3-component formalism, which is a mathematically transparent generalization of the additive augmentation given by Soler and Williams [11–13] (See discussion in Sec.VII in Ref.[6]). We give a formalism of the PMT method based on the 3-component formalism. In the PAW method [6], the total energy is minimized as a functional of pseudo wavefunctions. In the 3-component formalism, the minimization is formulated as for the wavefunctions represented in the *3-component space* defined in

Sec. II under some constraints. This is somehow general in the sense that it allows to use any kinds of basis (need not to be given by projectors); thus it is suitable to formulate mixed basis methods such as the PMT. Results of the PMT method applied to diatomic molecules from H_2 through Kr_2 are already given in Ref.14. Considering the fact that the PMT method (even the LMTO itself) is already pretty good to describe solids [5, 8, 15, 16], the PMT method can be a candidate to perform full-potential calculations for molecules and solids in an unified manner, more efficiently than LAPW.

Note that we had already implemented the quasiparticle self-consistent GW (QSGW) method [15, 17, 18] in the PMT method [19]; this allows us to apply the QSGW method to wide range of materials without empty spheres, with not being bothered with the difficulty of setting parameters of MTOs. We have applied the QSGW method to cases [20, 21]. This kind of method was referred to by Kimes, Katak and Kresse [22], who claimed that accurate and efficient GW calculations should be based on a method using both kinds of bases (localized and extended bases in space) simultaneously.

In Sec. II, we will give the 3-component formalism. Functional relations of physical quantities become transparent in the formalism. In Sec. III, we give the formulation of the PMT method based on the 3-component formalism. Then we discuss problems in the PMT method and ways to overcome them, giving a comparison with the PAW method. Derivation of atomic forces becomes straightforward as given in Appendix without any confusion that were discussed in Ref. [13].

II. 3-COMPONENTS FORMALISM

We assume periodic boundary condition where real space (or unit cell) is specified by Ω . Ω is divided into the muffin-tin (MT) regions and the interstitial region. The MTs are located at \mathbf{R}_a with radius R_a , where a is the index specifying a MT within Ω . $L \equiv (l, m)$ is the angular momentum index. We use units, $\hbar = 1$, electron mass m_e , and electron charge e . The spin index is not shown for simplicity.

Here we give the 3-component formalism as a general scheme for the augmented-wave methods, which include any kinds of augmented waves including the local orbitals [7].

A. the 3-component space

Any augmented basis $F_i(\mathbf{r})$ consists of three kinds of components, where i is the index specifying basis function. $F_i(\mathbf{r})$ consists of the following three components:

- (0) the smooth part (= envelope function) $F_{0i}(\mathbf{r})$
- (1) the true parts $F_{1i,a}(\mathbf{r})$ defined in MTs $|\mathbf{r}| \leq R_a$.

- (2) the counter parts $F_{2i,a}(\mathbf{r})$ defined in MTs $|\mathbf{r}| \leq R_a$ (canceling the smooth parts within MTs).

We call $F_{0i}(\mathbf{r})$, $F_{01,a}(\mathbf{r})$, and $F_{02,a}(\mathbf{r})$ as the 0th, 1st, and 2nd components of $F_i(\mathbf{r})$, respectively. The $F_{0i}(\mathbf{r})$ should be an analytic function in space or a linear combination of analytic functions. In the PMT method, $F_{0i}(\mathbf{r})$ are the PWs or the Bloch sums of the Hankel functions; exactly speaking, we use atom-centered smooth Hankel functions (smHankels) instead of the conventional Hankel functions, so as to avoid divergence at its center [5, 23] (See Eq. (37) and around). $F_{1i,a}(\mathbf{r})$ and $F_{2i,a}(\mathbf{r})$ are defined only at $|\mathbf{r}| \leq R_a$ (in cases below, we sometimes take these are zero at $|\mathbf{r}| > R_a$). In the sense that $F_{0i}(\mathbf{r})$ is analytic and 1st and 2nd components are given on a dense radial mesh, a basis $F_i(\mathbf{r})$ is specified without any numerical inaccuracy.

$F_i(\mathbf{r})$ is a member in the 3-component space, which is defined as a direct sum of linear spaces corresponding to the components (0), (1) and (2). Thus F_i can be expressed as $F_i = \{F_{0i}(\mathbf{r}), \{F_{1i,a}(\mathbf{r})\}, \{F_{2i,a}(\mathbf{r})\}\}$ (curly bracket mean a set), where $F_{1i} \equiv \{F_{1i,a}(\mathbf{r})\}$ means a set as for the MT index a , F_{2i} as well. However, in the followings, we use a little different expression instead:

$$\begin{aligned} F_i(\mathbf{r}) &= F_{0i}(\mathbf{r}) \oplus \{F_{1i,a}(\mathbf{r})\} \ominus \{F_{2i,a}(\mathbf{r})\} \\ &= F_{0i}(\mathbf{r}) \oplus F_{1i}(\mathbf{r}) \ominus F_{2i}(\mathbf{r}). \end{aligned} \quad (1)$$

This makes following expressions easy to read without any difference in their meanings. Symbols \oplus and \ominus mean nothing more than separators. We call a member in the 3-component space as a 3-component function in the followings. Wavefunctions are also given as 3-component functions. With the coefficients $\{\alpha_p^i\}$, wavefunctions can be written as

$$\psi_p(\mathbf{r}) = \sum_i \alpha_p^i F_i(\mathbf{r}), \quad (2)$$

where linear combinations are taken for each components. We represent electron density and so on as a 3-component function as well.

Note that the 3-component space is a mathematical construction, a model space: we have to specify how to map a 3-component function to a function in real space. For this purpose, we define \mathcal{A} -mapping (augmentation mapping) from a 3-component function to a function in real space;

$$\mathcal{A}[\psi_p(\mathbf{r})] \equiv \psi_{0p}(\mathbf{r}) + \sum_a \psi_{1p,a}(\mathbf{r} - \mathbf{R}_a) - \sum_a \psi_{2p,a}(\mathbf{r} - \mathbf{R}_a). \quad (3)$$

This is nothing but a conventional augmentation where physically meaningful wavefunctions $\psi_p(\mathbf{r})$ should satisfy following conditions (A) and (B);

- (A) Within MTs ($|\mathbf{r}| < R_a$), $\psi_{2p,a}(\mathbf{r}) = \psi_{0p}(\mathbf{r} + \mathbf{R}_a)$.
- (B) At MT boundaries ($|\mathbf{r}| = R_a$), $\psi_{1p,a}(\mathbf{r})$ and $\psi_{2p,a}(\mathbf{r})$ should have the same value and slope.

If (A) is satisfied, the contribution from ψ_{0p} within MTs perfectly cancels those of $\psi_{2p,a}$ in Eq. (3). The total energy in the DFT is given as a functional of eigenfunctions as $E[\{\psi_p(\mathbf{r})\}]$, where $\{\psi_p(\mathbf{r})\}$ are for occupied states. Our problem is to minimize this under the constraint of orthogonality of $\psi_p(\mathbf{r})$ with conditions (A) and (B) on $\{\psi_p(\mathbf{r})\}$. Local orbitals [7] is also treated as 3-component functions whose 0th and 2nd components are zero overall.

In the conventional LAPW (e.g. See [3, 4]), (A) and (B) are very accurately satisfied. The 2nd component almost completely satisfy (A) with the use of spherical Bessel functions. The 1st component are given up to very high l ($\gtrsim 8$). Thus the LAPW can be quite accurate. However, it can be expensive (we also have null-vector problem. See Sec. III A.).

Thus Soler and Williams [11] introduced additive augmentation: to make calculations efficient, we use condition (A') as a relaxed version of condition (A),

(A') Within MTs ($|\mathbf{r}| \leq R_a$), $\psi_{2p,a}(\mathbf{r}) \approx \psi_{0p}(\mathbf{r} + \mathbf{R}_a)$.

Then we expect high-energy (high frequency) contributions of eigenfunctions not included in the 1st and 2nd components are accounted for by the 0th component. In practice, we can use low l cutoff $\lesssim 4$ for both of 1st and 2nd components. A LAPW package HiLAPW, developed by Oguchi et al [24], implemented a procedure to evaluate physical quantities from the basis given by Eq. (3) with the condition (A').

However, it is complicated to evaluate all quasilocal products such as the density and kinetic-energy density from $\mathcal{A}[F_i^*(\mathbf{r})]\mathcal{A}[F_j(\mathbf{r}')]$, since it contains cross terms which connect different components. Thus Soler and Williams [11] gave a prescription to avoid the evaluation of the cross terms. With \mathcal{A} -mapping applied not to wavefunctions but to products of them as in Sec. II B, we have separable form of the total energy and all other physical quantities (no cross terms between components). This is based on the fact that the total energy in the separable form should agree with the true total energy only when

(A) and (B) are satisfied. As we see in the followings, it is a good approximation to use (A') instead of (A).

Above two important concepts, the additive augmentation and the separable form, were used in both of LMTO and PAW [5, 6, 10]. They were originally introduced in Ref.[11].

Let us consider how to determine $F_{1i,a}, F_{2i,a}$ for a given F_{0i} . As for $F_{2i,a}$, (A') means that F_{0i} should be reproduced well within MTs. Generally speaking, $F_{2i,a}(\mathbf{r})$ can be represented as

$$F_{2i,a}(\mathbf{r}) \equiv \sum_{k,L} C_{akL}^i P_{akL}(\mathbf{r}), \quad (4)$$

where k is index for radial degree of freedom. We introduce truncation parameters $k_{\max,a}$ and $l_{\max,a}$; we assume sum in Eq. (4) is taken for $k \leq k_{\max,a}$ and $l \leq l_{\max,a}$; when $k_{\max,a}$ and $l_{\max,a}$ becomes infinite, we assume condition (A) is satisfied. Even when these truncation parameters are finite, $F_{2i,a}$ should reproduce low energy (low frequency) parts of F_{0i} well. The functions $\{P_{akL}(\mathbf{r})\}$ can be rather general; as explained in Sec. III the central parts of smHankel is treated as it is (in other words, treated as a member of $\{P_{akL}(\mathbf{r})\}$ [25]). $F_{1i,a}(\mathbf{r})$ is given from Eq. (4) with a replacement of $P_{akL}(\mathbf{r})$ with $\tilde{P}_{akL}(\mathbf{r})$. Here $\tilde{P}_{akL}(\mathbf{r})$ is a linear combination of partial waves so as to have the same value and slope with $P_{akL}(\mathbf{r})$ at $|\mathbf{r}| = R_a$. With this replacement, we have

$$F_{1i,a}(\mathbf{r}) = \sum_{k,L} C_{akL}^i \tilde{P}_{akL}(\mathbf{r}). \quad (5)$$

B. augmentation for product of 3-component functions

Let us give a prescription to evaluate physical quantities for wavefunctions satisfying conditions (A') and (B). First, we define diagonal product of 3-component functions as

$$F_i^*(\mathbf{r})F_j(\mathbf{r}') \equiv F_{0i}^*(\mathbf{r})F_{0j}(\mathbf{r}') \oplus \{F_{1i,a}^*(\mathbf{r})F_{1j,a}(\mathbf{r}')\} \ominus \{F_{2i,a}^*(\mathbf{r})F_{2j,a}(\mathbf{r}')\}, \quad (6)$$

where we have no cross terms between different components. We apply \mathcal{A} -mapping in Eq. (3) to this product as

$$\begin{aligned} \mathcal{A}[F_i^*(\mathbf{r})F_j(\mathbf{r}')] &= \\ F_{0i}^*(\mathbf{r})F_{0j}(\mathbf{r}') &+ \sum_a F_{1i,a}^*(\mathbf{r}-\mathbf{R}_a)F_{1j,a}(\mathbf{r}'-\mathbf{R}_a) - \sum_a F_{2i,a}^*(\mathbf{r}-\mathbf{R}_a)F_{2j,a}(\mathbf{r}'-\mathbf{R}_a). \end{aligned} \quad (7)$$

We will use $\mathcal{A}[F_i^*(\mathbf{r})F_j(\mathbf{r}')]$ to evaluate quasilocal products when (A') is satisfied. Since any one-body quantities such as the inner product, electron density, current and so on, are quasilocal, we can evaluate these from $\mathcal{A}[\psi_p^*(\mathbf{r})\psi_{p'}(\mathbf{r}')]$. Generally speaking, we can evaluate matrix elements of a quasilocal operator $X(\mathbf{r}, \mathbf{r}')$ in real

space from 3-component wavefunctions $\psi_p(\mathbf{r})$ in separable form as

$$\langle \psi_p | X | \psi_{p'} \rangle = \int d^3r d^3r' X(\mathbf{r}, \mathbf{r}') \mathcal{A}[\psi_p^*(\mathbf{r})\psi_{p'}(\mathbf{r}')] \quad (8)$$

We can read this as a transformation of X to the corresponding operator in the 3-component space.

Based on the above prescription, we can define the inner product $\langle \psi_p | \psi_{p'} \rangle$ as $\langle \psi_p | \psi_{p'} \rangle = \sum_{i,j} \alpha_p^{i*} \alpha_{p'}^j O_{ij}$. Here the overlap matrix O_{ij} is:

$$O_{ij} \equiv \langle F_i | F_j \rangle \equiv \int_{\Omega} d^3r \mathcal{A}[F_i^*(\mathbf{r}) F_j(\mathbf{r})] \\ = \int_{\Omega} d^3r F_{0i}^*(\mathbf{r}) F_{0j}(\mathbf{r}) + \sum_a \int_{|\mathbf{r}| \leq R_a} d^3r F_{1i,a}^*(\mathbf{r}) F_{1j,a}(\mathbf{r}) - \sum_a \int_{|\mathbf{r}| \leq R_a} d^3r F_{2i,a}^*(\mathbf{r}) F_{2j,a}(\mathbf{r}). \quad (9)$$

This can read as a definition of the inner product in the 3-component space. For a given finite basis set, we can expect that O_{ij} should be positive definite as long as truncation parameters are large enough. The kinetic energy is given from $\rho_{ij} = \sum_p^{\text{occ.}} \alpha_p^{i*} \alpha_p^j$ (occ. means the sum for occupied states) as $E_k = \sum_{i,j} \rho_{ij} T_{ij}$. Here the kinetic-energy matrix T_{ij} is given as

$$T_{ij} \equiv \frac{\langle \nabla F_i | \nabla F_j \rangle}{2m_e} \equiv \frac{1}{2m_e} \int_{\Omega} d^3r (\nabla_{\mathbf{r}} \nabla_{\mathbf{r}'} \mathcal{A}[F_i^*(\mathbf{r}) F_j(\mathbf{r}')]_{\mathbf{r}=\mathbf{r}'} = \frac{1}{2m_e} \int_{\Omega} d^3r \mathcal{A}[\nabla F_i^*(\mathbf{r}) \nabla F_j(\mathbf{r})] \\ = \int_{\Omega} d^3r \frac{\nabla F_{0i}^*(\mathbf{r}) \nabla F_{0j}(\mathbf{r})}{2m_e} + \sum_a \int_{|\mathbf{r}| \leq R_a} d^3r \frac{\nabla F_{1i,a}^*(\mathbf{r}) \nabla F_{1j,a}(\mathbf{r})}{2m_e} - \sum_a \int_{|\mathbf{r}| \leq R_a} d^3r \frac{\nabla F_{2i,a}^*(\mathbf{r}) \nabla F_{2j,a}(\mathbf{r})}{2m_e}. \quad (10)$$

Partial integration gives $T_{ij} = \langle F_i | \frac{-\nabla^2}{2m_e} | F_j \rangle$, since $F_{1i,a}$ and $F_{2i,a}$ have the same value and slope at the MT boundaries.

This kinetic energy operator is interpreted as $T = \frac{-\nabla^2}{2m_e} \oplus \{ \frac{-\nabla^2}{2m_e} \} \ominus \{ \frac{-\nabla^2}{2m_e} \}$ in the 3-component space.

One-body problem for a given one-particle potential $V(\mathbf{r})$ in real space is translated into a problem in the 3-component space for the Hamiltonian $H = T + V$ under the condition (A) or (A'), where $V = V_0 \oplus \{V_{1,a}\} \ominus \{V_{2,a}\}$. Here $V_0(\mathbf{r}) = V(\mathbf{r})$, and $V_{1,a}(\mathbf{r}) = V_{2,a}(\mathbf{r}) = V(\mathbf{r} + \mathbf{R}_a)$ within MTs at \mathbf{R}_a . However, we can add any extra potential $\Delta \bar{V}$ simultaneously to both of V_0 and $V_{2,a}$ if (A) is completely satisfied.

We have an error because we use Eq. (7) instead of Eq. (3): high energy contributions contained in the 0th components are not exactly evaluated. However, the error can be small enough to be neglected as discussed in Appendix A. This error is also related to a question, how to choose the optimum $\Delta \bar{V}$ so as to minimize the error. In fact, the success of the PAW [6] is dependent on the choice of $\Delta \bar{V}$ as seen in Sec. III B.

The valence electron density n as the 3-component function is given by

$$n = n_0 \oplus n_1 \ominus n_2 = n_0 \oplus \{n_{1,a}\} \ominus \{n_{2,a}\} = \sum_{ij} \rho_{ij} F_i^* F_j = \sum_{ij} \rho_{ij} F_{0i}^*(\mathbf{r}) F_{0j}(\mathbf{r}) \\ \oplus \{ \sum_{ij} \rho_{ij} F_{1i,a}^*(\mathbf{r}) F_{1j,a}(\mathbf{r}) \} \ominus \{ \sum_{ij} \rho_{ij} F_{2i,a}^*(\mathbf{r}) F_{2j,a}(\mathbf{r}) \}. \quad (11)$$

We can calculate the Coulomb interaction from $\mathcal{A}[n]$. However, to reduce the computational effort, we will also make the Coulomb interaction into the separable form as seen in Sec. II D, with the help of multipole technique due to Weinert [26]. In Sec. II C and Sec. II E, we give some preparations to define the Coulomb interaction in Sec. II D.

The total energy should be given as a functional of eigenfunctions in the first-principle calculations, not just as a functional of coefficients $\{\alpha_p^j\}$. This is important in some cases. For example, it is necessary to know how the change in the basis set affects the total energy when we calculate atomic forces. These are related to the so-called Pulay terms [27].

C. multipole transformation

In order to define Coulomb interaction in Sec. II D, we introduce the multipole transformation (\mathcal{M} -transformation) for the 3-component functions. This corresponds to the compensation charges in Ref. [6].

Before defining the \mathcal{M} -transformation, we define the gaussian projection $\mathcal{G}_a[f(\mathbf{r})]$ as follows. The projection $\mathcal{G}_a[f(\mathbf{r})]$ is defined for the function $f(\mathbf{r})$ for $|\mathbf{r}| \leq R_a$ as

$$\mathcal{G}_a[f(\mathbf{r})] = \sum_L Q_{aL}[f] G_{aL}(\mathbf{r}), \quad (12)$$

$$G_{aL}(\mathbf{r}) = \frac{1}{N_{aL}} \exp\left(-\left(\frac{r}{R_{G,a}}\right)^2\right) Y_L(\hat{\mathbf{r}}), \quad (13)$$

where $Q_{aL}[f] = \int_{|\mathbf{r}| \leq R_a} \mathcal{Y}_L(\mathbf{r}) f(\mathbf{r}) d^3r$ gives the L -th multipole moment of $f(\mathbf{r})$. Here $\mathcal{Y}_L(\mathbf{r}) \equiv r^L Y_L(\hat{\mathbf{r}})$. $Y_L(\hat{\mathbf{r}})$ is

the real spherical harmonics, where $\hat{\mathbf{r}}$ is the normalized \mathbf{r} . N_{aL} is a normalization factor so that $G_{aL}(\mathbf{r})$ has a normalized multipole moment. $R_{G,a}$ in Eq.(13) is chosen small enough so that $G_{aL}(\mathbf{r})$ is negligible for $|\mathbf{r}| \geq R_a$ (See Eq.(25) in Ref.[5]). This $\mathcal{G}_a[f(\mathbf{r})]$ is a superposition of gaussians $G_{aL}(\mathbf{r})$ with keeping the multipole moments of $f(\mathbf{r})$. We can take rather small $R_{G,a}$ without

loss of numerical accuracy; it is possible to take a limit $R_{G,a} \rightarrow 0$ because quantities involved in $G_{aL}(\mathbf{r})$ are evaluated analytically or numerically accurately on a dense radial mesh.

We now define \mathcal{M} -transformation for 3-component density $n = n_0 \oplus n_1 \oplus n_2$ as

$$\begin{aligned} \mathcal{M}[n] &= n_0(\mathbf{r}) \\ &+ \sum_{a,\mathbf{T},L} Q_{aL}[n_{1,a}-n_{2,a}]G_{aL}(\mathbf{r}-\mathbf{R}_a-\mathbf{T}) \oplus n_1 \ominus \{n_{2,a}(\mathbf{r}) + \sum_L Q_{aL}[n_{1,a}-n_{2,a}]G_{aL}(\mathbf{r})\}. \end{aligned} \quad (14)$$

Thus $\mathcal{M}[n]$ adds the same gaussians to both of the 0th and 2nd components. \mathbf{T} is the translational vectors of Ω . With this transformation, the multipole moments of the 1st and 2nd components become the same. Note that the \mathcal{M} -transformation is not a physically meaningful transformation because $\mathcal{A}[\mathcal{M}[n]] = \mathcal{A}[n]$. With this transformation, interstitial electrostatic potential calculated from the 0th component of Eq. (14) should be the

same as that calculated from $\mathcal{A}[n]$.

D. Coulomb interaction

In principle, we can define the Coulomb interaction between $n(\mathbf{r}) = n_0 \oplus n_1 \oplus n_2$ and $m(\mathbf{r}) = m_0 \oplus m_1 \oplus m_2$ from the densities $\mathcal{A}[n]$ and $\mathcal{A}[m]$. We can use $\mathcal{A}[\bar{n}]$ instead of $\mathcal{A}[n]$ where $\bar{n} = \mathcal{M}[n]$ satisfies $\mathcal{A}[\bar{n}] = \mathcal{A}[n]$, and $\mathcal{A}[\bar{m}]$ as well. Thus the Coulomb interaction $(n|v|m)_{\text{original}}$ is given as

$$(n|v|m)_{\text{original}} = \sum_{\mathbf{T}} \int_{\Omega} d^3r d^3r' \mathcal{A}[\bar{n}(\mathbf{r})] v(\mathbf{r}-\mathbf{r}'+\mathbf{T}) \mathcal{A}[\bar{m}(\mathbf{r}')]. \quad (15)$$

Here $v(\mathbf{r}) = e^2/|\mathbf{r}|$; $\sum_{\mathbf{T}}$ implicitly includes the division by number of cells. Equation (15) can not be easily evaluated because $v(\mathbf{r}-\mathbf{r}'+\mathbf{T})$ contains the cross terms which connect the 0th component with other components.

Thus we use an approximation

$$(n|v|m) \equiv \mathcal{M}[n] \cdot v \cdot \mathcal{M}[m] = \bar{n} \cdot v \cdot \bar{m}, \quad (16)$$

instead of Eq. (15), where dot operator for the 3-component functions is given as

$$\bar{n} \cdot v \cdot \bar{m} \equiv \bar{n}_0 \bullet v \bullet \bar{m}_0 + \bar{n}_1 \circ v \circ \bar{m}_1 - \bar{n}_2 \circ v \circ \bar{m}_2 \quad (17)$$

$$\bar{n}_0 \bullet v \bullet \bar{m}_0 \equiv \sum_{\mathbf{T}} \int_{\Omega} d^3r d^3r' \bar{n}_0(\mathbf{r}) v(\mathbf{r}-\mathbf{r}'+\mathbf{T}) \bar{m}_0(\mathbf{r}'), \quad (18)$$

$$\bar{n}_1 \circ v \circ \bar{m}_1 \equiv \sum_a \int_{|\mathbf{r}| \leq R_a} d^3r \int_{|\mathbf{r}'| \leq R_a} d^3r' \bar{n}_{1,a}(\mathbf{r}) v(\mathbf{r}-\mathbf{r}') \bar{m}_{1,a}(\mathbf{r}'), \quad (19)$$

$$\bar{n}_2 \circ v \circ \bar{m}_2 \equiv \sum_a \int_{|\mathbf{r}| \leq R_a} d^3r \int_{|\mathbf{r}'| \leq R_a} d^3r' \bar{n}_{2,a}(\mathbf{r}) v(\mathbf{r}-\mathbf{r}') \bar{m}_{2,a}(\mathbf{r}'). \quad (20)$$

Note that $X \bullet Y$ means integral over Ω , whereas $X \circ Y$ means integrals within MTs.

Let us evaluate the difference between Eq. (15) and Eq. (16). This can be evaluated with the identity in Appendix A as

$$\begin{aligned} (n|v|m)_{\text{original}} - (n|v|m) &= \sum_a \int_{|\mathbf{r}| \leq R_a} d^3r \int_{|\mathbf{r}'| \leq R_a} d^3r' \left((\bar{n}_0(\mathbf{r}) - \bar{n}_2(\mathbf{r})) v(\mathbf{r}-\mathbf{r}') (\bar{m}_1(\mathbf{r}') - \bar{m}_2(\mathbf{r}')) \right. \\ &\quad \left. + (\bar{n}_1(\mathbf{r}) - \bar{n}_2(\mathbf{r})) v(\mathbf{r}-\mathbf{r}') (\bar{m}_0(\mathbf{r}') - \bar{m}_2(\mathbf{r}')) \right). \end{aligned} \quad (21)$$

This is essentially the same with Eq.(13) in Ref.[10]. In Eq. (21), the difference consists of contributions from MT sites without terms connecting different MT sites. This is because $\bar{n}_{1,a}(\mathbf{r})$ and $\bar{n}_{2,a}(\mathbf{r})$ have the same multipole moments.

Since $\bar{n}_0(\mathbf{r}') - \bar{n}_2(\mathbf{r})$ is high- l or highly oscillating part, and $\bar{n}_{1,a}(\mathbf{r}) - \bar{n}_{2,a}(\mathbf{r})$ has zero multipole moments and zero at MT boundaries, we expect that the separable form of Eq. (16) should be justified. We can check this with changing the truncation parameters $l_{\max,a}$ and $k_{\max,a}$.

From Eq. (16), we have the expression of the Coulomb interaction as

$$(F_i^* F_j | v | F_{i'}^* F_{j'}) = \mathcal{M}[F_i^* F_j] \cdot v \cdot \mathcal{M}[F_{i'}^* F_{j'}]. \quad (22)$$

Here $F_i^* F_j$ is the diagonal product defined in Eq. (6) at $\mathbf{r} = \mathbf{r}'$. In calculations such as arising in the GW approximations [15], we have to evaluate this as accurately as possible so that the exchange-pair cancellation is kept well.

E. Frozen core approximation

We often need to treat spillout of the core density outside of MTs explicitly. Then we use the frozen core approximation; the charge density due to the cores are evaluated by a superposition of rigid cores as follows [5].

First, we perform a self-consistent atomic calculation under the spherical approximation without a spin polarization to obtain its core density $n_a^c(\mathbf{r})$. Then we make a fitted density $n_{\text{SH},a}^c(\mathbf{r})$ given by a linear combination of several smHankel functions so that $n_{\text{SH},a}^c(\mathbf{r})$ reproduces $n_a^c(\mathbf{r})$ for $|\mathbf{r}| > R_a$ within a numerical accuracy. Since $n_{\text{H},a}^c(\mathbf{r})$ are analytic and smooth at their centers, we can treat them numerically accurately (we can use other kinds of analytic functions such as gaussians instead of smHankel functions).

Thus we have the expression of all the core electron density with adding contribution from nucleus $-Z_a\delta(\mathbf{r})$:

$$n^{\text{Zc}} = \sum_{a,\mathbf{T}} n_{\text{SH},a}^c(\mathbf{r} - \mathbf{R}_a - \mathbf{T}) \oplus \{n_a^c(\mathbf{r}) - Z_a\delta(\mathbf{r})\} \ominus \{n_{\text{SH},a}^c(\mathbf{r})\}. \quad (23)$$

Applying the \mathcal{M} -transformation to n^{Zc} gives

$$\begin{aligned} \mathcal{M}[n^{\text{Zc}}] &= \sum_{a,\mathbf{T}} \left(n_{\text{SH},a}^c(\mathbf{r} - \mathbf{R}_a - \mathbf{T}) + \sum_L Q_{aL}^{\text{Zc}} G_{aL}(\mathbf{r} - \mathbf{R}_a - \mathbf{T}) \right) \oplus \{n_a^c(\mathbf{r}) - Z_a\delta(\mathbf{r})\} \\ &\ominus \{n_{\text{SH},a}^c(\mathbf{r}) + \sum_L Q_{aL}^{\text{Zc}} G_{aL}(\mathbf{r})\}, \end{aligned} \quad (24)$$

$$Q_{aL}^{\text{Zc}} = Q_{aL}[n_1^{\text{Zc}} - n_2^{\text{Zc}}] = Q_{aL}[n_a^c(\mathbf{r}) - Z_a\delta(\mathbf{r}) - n_{\text{SH},a}^c(\mathbf{r})]. \quad (25)$$

F. total energy in density functional

Let us give the total energy E_{total} for the DFT, and construct the Kohn-Sham equation from it. With the kinetic energy $E_k = \frac{1}{2m_e} \sum_{ij} \rho_{ij} \langle \nabla F_i | \nabla F_j \rangle$ from Eq.(10), the total energy is given as:

$$E_{\text{total}} = E_k^{\text{core}} + E_k + E_{\text{es}} + E_{\text{xc}}, \quad (26)$$

where E_k^{core} is the kinetic energy of frozen cores as a constant. E_{es} and E_{xc} are electrostatic and exchange-correlation energies, respectively. E_{es} is given as the electrostatic energy for the total density $n^{\text{Zcv}} = n^{\text{Zc}} + n$, which are given in Eqs(11,23).

Based on the definition Eq. (22), we have

$$E_{\text{es}} = \frac{1}{2} (n^{\text{Zcv}} | v | n^{\text{Zcv}}) = \frac{1}{2} \mathcal{M}[n^{\text{Zcv}}] \cdot v \cdot \mathcal{M}[n^{\text{Zcv}}], \quad (27)$$

where a constant due to the self-interaction of nucleus is implicitly removed. Components of $\bar{n}^{\text{Zcv}}(\mathbf{r}) = \mathcal{M}[n^{\text{Zcv}}]$ are given as

$$\bar{n}_0^{\text{Zcv}}(\mathbf{r}) = n_0^{\text{Zc}}(\mathbf{r}) + \sum_{a,L,\mathbf{T}} (Q_{aL}^{\text{Zc}} + Q_{aL}^{\text{v}}) G_{aL}(\mathbf{r} - \mathbf{R}_a - \mathbf{T}) + n_0(\mathbf{r}), \quad (28)$$

$$\bar{n}_{1,a}^{\text{Zcv}}(\mathbf{r}) = n_{1,a}^{\text{Zc}}(\mathbf{r}) + n_{1,a}(\mathbf{r}), \quad (29)$$

$$\bar{n}_{2,a}^{\text{Zcv}}(\mathbf{r}) = n_{2,a}^{\text{Zc}}(\mathbf{r}) + \sum_L (Q_{aL}^{\text{Zc}} + Q_{aL}^{\text{v}}) G_{aL}(\mathbf{r}) + n_{2,a}(\mathbf{r}), \quad (30)$$

where $Q_{aL}^v = Q_{aL}[n_{1,a} - n_{2,a}]$. We expand $F_{0i}^*(\mathbf{r})F_{0j}(\mathbf{r})$ of n_0 in $\{e^{i\mathbf{G}\mathbf{r}}\}$ (to obtain coefficients, $F_{0i}^*(\mathbf{r})F_{0j}(\mathbf{r})$ is tabulated on a real-space mesh, then it is Fourier transformed). The cutoff on \mathbf{G} is specified by $E_{\text{MAX}}^{\text{mesh}}$. Then the 0th components in Eq. (28) is represented by sum of analytic functions. Thus we can finally calculate $\frac{1}{2}\bar{n}_0^{\text{Zcv}}(\mathbf{r}) \bullet v \bullet \bar{n}_0^{\text{Zcv}}(\mathbf{r})$ in E_{es} . Terms between gaussians located at different MT sites are evaluated with the Ewald sum treatment. The terms related to MTs in E_{es} is $\frac{1}{2}\bar{n}_1^{\text{Zcv}} \circ w \circ \bar{n}_1^{\text{Zcv}} - \frac{1}{2}\bar{n}_2^{\text{Zcv}} \circ w \circ \bar{n}_2^{\text{Zcv}}$, which is calculated on a radial mesh accurately.

The exchange correlation term can be defined as

$$E_{\text{xc}}[n^{\text{Zcv}}] = E_{\text{xc}}[n_0^{\text{Zcv}}] + \sum_a E_{\text{xc}}[n_{1,a}^{\text{Zcv}}] - \sum_a E_{\text{xc}}[n_{2,a}^{\text{Zcv}}]. \quad (31)$$

The functional derivatives of $E_{\text{xc}}[n^{\text{Zcv}}]$ with respect to each component of n^{Zcv} gives

$$v^{\text{xc}} = v_0^{\text{xc}}(\mathbf{r}) \oplus \{v_{1,a}^{\text{xc}}(\mathbf{r})\} \ominus \{v_{2,a}^{\text{xc}}(\mathbf{r})\}. \quad (32)$$

To determine the ground state, E_{total} should be minimized under the orthogonality of eigenfunctions with the constraint (A') and (B). This ends up with $\delta\psi_p^* \cdot (H - \epsilon_p) \cdot \psi_p = 0$ for the variation $\delta\psi_p^*$ which satisfy (A') and (B). Here the operator $H = T + V$ is given as

$$T = \frac{-\nabla^2}{2m_e} \oplus \left\{ \frac{-\nabla^2}{2m_e} \right\} \ominus \left\{ \frac{-\nabla^2}{2m_e} \right\} \quad (33)$$

$$V = \bar{n}_0^{\text{Zcv}} \bullet v \bullet v_0^{\text{xc}} \oplus \left\{ \sum_L Q_{aL}^v \mathcal{Y}_L(\mathbf{r}) + \bar{n}_{1,a}^{\text{Zcv}} \circ w \circ v_{1,a}^{\text{xc}} \right\} \ominus \left\{ \sum_L Q_{aL}^v \mathcal{Y}_L(\mathbf{r}) + \bar{n}_{2,a}^{\text{Zcv}} \circ w \circ v_{2,a}^{\text{xc}} \right\}, \quad (34)$$

$$Q_{aL}^v \equiv \frac{\partial E_{\text{es}}}{\partial Q_{aL}^v} = \bar{n}_0^{\text{Zcv}} \bullet v \bullet G_{aL}(\mathbf{r}' - \mathbf{R}_a) - \bar{n}_{2,a}^{\text{Zcv}} \circ w \circ G_{aL}(\mathbf{r}'), \quad (35)$$

where $\bar{n}_0 \bullet v \bullet$ means an integral on a variable, resulting a function of \mathbf{r} .

When a basis set $\{F_j(\mathbf{r})\}$ satisfying (A') and (B) are fixed, we just need to consider variation with respect to α_p^{i*} in Eq. (2). Then we have

$$\sum_j (H_{ij} - \epsilon_p O_{ij}) \alpha_p^j = 0, \quad (36)$$

where $H_{ij} = \langle F_i | H | F_j \rangle = \langle F_i | \frac{-\Delta}{2m} + V | F_j \rangle = T_{ij} + V_{ij}$. $V_{ij} = \langle F_i | V | F_j \rangle = V \cdot F_i^* F_j$. Then the total energy minimization results in the eigenvalue problem. The matrix elements O_{ij} , T_{ij} and V_{ij} are given in Appendix C.

The formula to evaluate atomic forces are given in Appendix B. It is directly evaluated from the variation on the total energy. This procedure is considerably simplified than that given in Refs.[5, 9].

III. PMT METHOD

Let us give the PMT method based on the 3-component formalism in Sec. II. Based on it, we need to specify a basis set $\{F_i\}$. In the PMT, $\{F_i\}$ is classified into three kinds of subsets as follows:

- (a) APW. We augment the PW in the manner as will be shown later.
- (b) MTO. We augment the atom-centered smHankel functions.
- (c) Local orbital (Lo) [28]. We use this to represent some degree of freedom in MTs, such as semicore states. The envelope function of Lo is zero overall.

The smHankel function, as the envelop function of MTO, is first introduced by Methfessel [5, 23]. The spherical smHankel function $h_0(\mathbf{r})$ (for $l = 0$) is defined by the Helmholtz equation with a gaussian source term $g_0(\mathbf{r}) = C \exp(-r^2/R_{\text{SM}}^2)$ (see Eq.(5) in Ref.[5]) instead of δ -function;

$$(\nabla^2 + \epsilon)h_0(\mathbf{r}) = -4\pi g_0(\mathbf{r}), \quad (37)$$

where $C = 1/(\sqrt{\pi}R_{\text{SM}})^3$ is the normalization constant. $\epsilon = -\kappa^2$ is the negative energy to specify the asymptotic damping behavior of $h_0(\mathbf{r})$. At the limit $R_{\text{SM}} \rightarrow 0$ where $g_0(\mathbf{r})$ becomes δ -function (as a point charge), $h_0(\mathbf{r})$ becomes to the Hankel function $h_0(\mathbf{r}) = \exp(-\kappa r)/r$. Since the source term is smeared with the radius R_{SM} , we have no divergent behavior at $r = 0$ anymore; the smHankel bends over at $\sim R_{\text{SM}}$ (See Fig.1 in Ref.[5]). From $h_0(\mathbf{r})$, we can make $h_L(\mathbf{r}) \equiv \mathcal{Y}_L(-\nabla)h_0(\mathbf{r})$ for any L . (recall $\mathcal{Y}_L(\mathbf{r}) = r^L Y_L(\hat{\mathbf{r}})$). $\mathcal{Y}_L(-\nabla)$ means to substitute \mathbf{r} in $\mathcal{Y}_L(\mathbf{r})$ with $-\nabla$. See Ref.[23] for details.

For the augmentation of the PW, that is, to determine the 2nd component from PW as 0th component, we expand the PW within the MTs into the Laguerre polynomial [8]. Any function $f(\mathbf{r})$ (PW in this case) is

expanded within a MT $|\mathbf{r} - \mathbf{R}_a| \leq R_a$ as

$$f(\mathbf{r}) = \sum_{k,l} C_{akL}[f] P_{akL}(\mathbf{r} - \mathbf{R}_a), \quad (38)$$

$$P_{akL}(\mathbf{r}) = p_{akl}(r) Y_L(\hat{\mathbf{r}}), \quad (39)$$

where $k = 0, 1, 2, \dots$ denotes the order of a polynomials $p_{akl}(r)$. In the case that $f(\mathbf{r})$ is a PW, the coefficients for the function $C_{akL}[f]$ are given analytically [23].

When we use smHankel centered at \mathbf{R}_a as an envelope function $f(\mathbf{r})$, we have head part, which is $f(\mathbf{r}) = h_L(\mathbf{r} - \mathbf{R}_a)$ for $|\mathbf{r} - \mathbf{R}_a| \leq R_a$, and tail part, which is in other MT sites $|\mathbf{r} - \mathbf{R}_{a'}| \leq R_{a'}$. As for the tail part, we use the expansion of Eq.(38) as in the case of PW. On the other hand, we use the head part as it is [25]; this can be taken into account in the formalism if the set $\{P_{akL}(\mathbf{r})\}$ contains not only the Laguerre polynomials but also $h_L(\mathbf{r})$ as its members.

After specifying $\{P_{akL}(\mathbf{r})\}$, we can determine corresponding $\{\tilde{P}_{akL}(\mathbf{r})\}$ as a linear combination of $\phi_{al}(r) Y_L(\hat{\mathbf{r}})$ and $\dot{\phi}_{al}(r) Y_L(\hat{\mathbf{r}})$, where partial waves $\phi_{al}(r)$ and its energy derivatives $\dot{\phi}_{al}(r)$ are given as the solutions of the radial Schrödinger equation for the spherically-averaged potential of $V_{1,a}$ in Eq. (34), where energies E_{al} to solve the equation are given as the center of gravities of the occupied states of the partial density of states of the al component; thus $\phi(r)$ and $\dot{\phi}(r)$ are not with the subscripts aL but with al . This prescription to determine $\{\tilde{P}_{akL}(\mathbf{r})\}$ can be taken as a quasi-minimization procedure, from the view of total energy minimization.

As for the al with Lo, we need another partial wave $\phi_{al}^{\text{Lo}}(r)$ corresponding to Lo. When the Lo is to describe a deeper level, we can set the energy to solve the radial Schrödinger equation E_{al}^{Lo} at the center of gravity; then we set E_{al} at the Fermi energy instead of the prescription in the previous paragraph.

The number of basis is simply specified by the cutoff energy of the APW for (a). However, specification of MTOs (b) is not so simple. We use multiple MTOs for each aL to reduce the number of basis with keeping the computational accuracy [8]. Since $h_L(\mathbf{r})$ as the envelope functions are specified by the parameters R_{SM} and ϵ , we have to specify them for all MTOs. Ref.[5] discussed optimization of them so as to minimize the total energy. However, as seen in figures in Ref.[5], such non-linear optimization is too complicated. Thus it is necessary to give a method to set the parameters in a simple manner as follows. As for R_{SM} , we can use a condition $R_{\text{SM}} = R_a/2$ for all MTOs. Then the envelope functions out side of MTs well coincide with the usual Hankel function. Even with this simple setting of R_{SM} without optimization, numerical accuracy can be kept well; we can check the convergence of calculations with the number of APWs. We also see the dependence on ϵ 's are rather small in the PMT method. The dependence becomes less when we use larger number of APWs; hence we do not need to stick to careful choice of the parameter ϵ . Thus the serious problem of the full-potential LMTO method, "how

to choose MTO parameters" are essentially removed in the PMT method. This is numerically detailed in the paper which gives results for diatomic molecules from H_2 through Kr_2 [14].

We use one further approximation. In Eq.(11), we make angular-momentum cutoff. Even though we have angular momentum component up to $2 \times l_{\text{max},a}$ in the 1st and 2nd components in Eq.(11), we drop components higher than $l_{\text{max},a}$; it is meaningless to take them into account since we have already make truncations for eigenfunctions. Note that this does not affects O_{ij} and T_{ij} because only the special components determine them.

A. problems in the PMT method

Let us examine three problems of the PMT methods, and ways to manage them.

The first problem is the positive definiteness of O_{ij} . Since the last term in Eq.(9) can give negative contribution, there is a possibility that O_{ij} can not be positive definite. In principle, we can expect almost zero eigenvalues on the matrix $\int_{|\mathbf{r}| \leq R_a} d^3r (F_{0i}^*(\mathbf{r}) F_{0j}(\mathbf{r}) - F_{2i,a}^*(\mathbf{r}) F_{2j,a}(\mathbf{r}))$ for all MTs if the truncation parameters are large enough. This guarantees the positive definiteness of O_{ij} . In practice, we typically use $k_{\text{max},a} \sim 5$ and $l_{\text{max},a} \sim 4$; they can give satisfactory results with keeping positive definiteness of O_{ij} , as seen in Refs.[8, 14].

The second is the undefiniteness of the second component ψ_{2p} . This is clear if (A) is satisfied; as ψ_{2p} within MTs is not uniquely determined since it is canceled completely by ψ_{0p} within MTs. However, since we use (A') in practice, this can cause numerical instability. To illustrate this, let us consider a linear combination of basis functions where only their 0th and 2nd components within MT are non zero. This is a null vector which has no physical meanings; it gives zero when we apply Hamiltonian and Overlap matrix to it. This is a kind of ghost. Apparently, this occurs because the 3-component space is not a complete metric space in the mathematical sense. When we enlarge number of basis, this null vector can cause numerical problems. It can be an origin of uncontrollable eigenvalue (e.g, 0 divided by 0), or it can attach to some eigenfunctions and deform them easily. In fact, we observed unconverged cases when the 2nd component of electron density becomes too large. Within our current implementation of the PMT, we should use limited number of basis so as to avoid this problem. However, in Refs.[8, 14], we can see enough stability on the total energy convergence before such problems occurs when we increase the number of basis.

It will be possible to remove such undefiniteness in some manners. For example, we can minimize the total energy with adding a fixing term $+\lambda \sum_p \int_{|\mathbf{r}| \leq R_a} d^3r \psi_{2p,a}^*(\mathbf{r}) (1 - \tilde{P}) \psi_{2p,a}(\mathbf{r})$, where λ is a Lagrange multiplier, \tilde{P} is a projector to the space spanned by some pseudo partial waves corresponding to

true atomic partial waves. If λ is infinite, 2nd components are only spanned by the pseudo partial waves. However, we should avoid a large λ so as not to deteriorate the total energy minimization.

The third problem is the orthogonality to the cores. In the frozen core approximation in Sec. II E, we take account of the spillout of the core electron density from MTs; this allows us to use a small MT radius. However, when we use quite small MTs, we observed a problem of orthogonality of wavefunctions to the cores, resulting in unconvergence. In such a case, we need to introduce local orbitals to represent cores so as to keep the orthogonality. It may be possible to enforce the orthogonality with a projector as described in Ref.[6].

B. comparison with PAW

Here we will make a comparison of the PMT method with the PAW method [6, 10] based on the 3-component formalism.

In the PAW method, we perform the all-electron (AE) calculations for a spherical atom as a reference in advance. Then the main problem is how to solve the one-body problem for a given one-body potential $V(\mathbf{r})$ in real space. As in Sec. II B, the problem is translated into the problem in the 3-component space for $V = V_0 \oplus V_1 \ominus V_2$. For simplicity, we omit the index a in the followings.

The basis set in the PAW is given as follows. We first prepare AE partial waves $\{\phi_i(\mathbf{r})\}$ (e.g, two for each aL in Ref.[10]), as solutions of radial Schödinger eq. for V_1 at some reference energies $\{\epsilon_i\}$ (in this section, the index i is for the partial wave). Then we set up corresponding pseudo partial waves $\{\tilde{\phi}_i(\mathbf{r})\}$. The eigenfunction ψ in the PAW can be represented in the 3-component space; for given 0th-component $\bar{\psi}$ (this is called as *pseudo wavefunction*), we have ψ with projectors $\{\tilde{p}_i\}$ as

$$\psi = \bar{\psi} \oplus \sum_i |\phi_i\rangle \langle \tilde{p}_i | \bar{\psi} \rangle \ominus \sum_i |\tilde{\phi}_i\rangle \langle \tilde{p}_i | \bar{\psi} \rangle. \quad (40)$$

Here \tilde{p}_i should satisfy $\langle \tilde{p}_i | \tilde{\phi}_j \rangle = \delta_{ij}$. The minimization of the total energy of the one-body problem $E = \sum_j^{\text{occupied}} \psi_j^* \cdot (T + V) \cdot \psi_j$ with respect to $\bar{\psi}_j$ is given by

$$\left(\frac{-\nabla^2}{2m} + V_0(\mathbf{r}) - \epsilon_j + \sum_{ii'} |\tilde{p}_i\rangle (dH_{ii'} - \epsilon_j dO_{ii'}) \langle \tilde{p}_i| \right) \bar{\psi}_j = 0, \quad (41)$$

$$dH_{ii'} = \langle \phi_i | \frac{-\nabla^2}{2m} + V_1 | \phi_{i'} \rangle - \langle \tilde{\phi}_i | \frac{-\nabla^2}{2m} + V_2 | \tilde{\phi}_{i'} \rangle \quad (42)$$

$$dO_{ii'} = \langle \phi_i | \phi_{i'} \rangle - \langle \tilde{\phi}_i | \tilde{\phi}_{i'} \rangle. \quad (43)$$

If we use infinite number of partial waves which makes a complete set, Eq.(41) reproduces the original one-body problem in real space.

Let us consider a case where $\psi_j = \bar{\psi}_j \oplus \phi_j \ominus \tilde{\phi}_j$ is the solution of Eq. (41) with eigenvalue ϵ_j , where $\bar{\psi}_j$ within MT coincides with $\tilde{\phi}_j$. This is given by Eq. (40) from $\bar{\psi}_j$. When we make a truncation for the number of partial waves, $\{\tilde{p}_i\}$ should satisfy

$$\left(\frac{-\nabla^2}{2m} + V_0(\mathbf{r}) - \epsilon_j \right) |\tilde{\phi}_j\rangle + \sum_i |\tilde{p}_i\rangle (dH_{ij} - \epsilon_j dO_{ij}) = 0, \quad (44)$$

in order to satisfy Eq. (41). This determines $\{\tilde{p}_i\}$; this is one of the main idea in the PAW method. In practice, considering the numerical stability, we determine \tilde{p}_i so that Eq. (44) is approximately satisfied [6].

Another important idea of the PAW is the introduction of the pseudopotential. This is how to determine V_0 within MT ($= V_2$). This is because the result strongly depends on the pseudopotential when the number of partial waves are small. In principle, the pseudopotential should be determined so that $\bar{\psi}_j$ contain high energy part (high angular momentum l or highly oscillating part) of the wavefunctions which is missing in the 1st and 2nd components due to the truncation of the number of partial waves.

Note that the truncation can cause the ghost state problem in the PAW method. To illustrate this, consider a case that s wave part in MT is described only by two partial waves $2s$ and $3s$. Then the PAW procedure maps $\bar{\psi}$ with zero node to ψ with one node, $\bar{\psi}$ with one node to ψ with two nodes. Problem is that $\bar{\psi}$ with two nodes, which is orthogonal to $\{\tilde{\psi}_i\}$ for $2s$ and $3s$, can not be mapped to ψ with three nodes due to the truncation. Thus it is possible that such a function cause a ghost state; we have to design the pseudopotential so that such $\bar{\psi}$ should be kept to be at a high enough energy region (to push $\bar{\psi}$ high away from the Fermi energy, it may be better to use relatively strongly repulsive pseudopotential). Ref.[10] claims that there is no ghost state for all kinds of

atoms. However, it is not easy to check the convergence within the framework of the PAW method.

In the PAW method with PWs proposed in Ref.[10], many PWs are required compared with with LAPW. Roughly speaking, energy cutoff of PWs are $\sim 15\text{Ry}$ in LAPW, and $\sim 30\text{Ry}$ in PAW [10, 29]. This is because the PAW method, as is the case of pseudopotential methods, needs to uniquely determine the pseudo partial waves (0th component) within MT. This is in contrast with the LAPW (and the PMT) method, where 0th component within MT is irrelevant because the 2nd components have enough degree of freedom to well cancel its contribution. However, with sacrificing the cutoff energy, the PAW takes robust convergence that comes from the absence of the null vector problem discussed in Sec.III A

As a theoretical possibility, we can imagine a method to use smHankels together with the PWs in the basis set for the one-body problem in the PAW method. However, it is not very clear whether it becomes a efficient method or not. To reduce the number of basis of PWs, it is necessary to make the smHankels span high-energy parts of pseudo wavefunctions. Thus we have to tailor smHankel so that it fits to the pseudo wavefunctions not

only interstitial region, but also within MT. This can be not straightforward.

IV. SUMMARY

We have reformulated the PMT method on the basis of the 3-component formalism, which is a generalized version of the additive augmentation given by Soler and Williams. The 3-component formalism allows including any kinds of basis not necessarily given by a projector as PAW. This fits the procedure to give the Kohn-Sham equation for a mixed basis method such as the PMT method from the total energy minimization scheme; this results in the transparent derivation of the atomic forces. We believe that the formalism shown here could give a basis for future developments. Our results for molecules from H_2 through Kr_2 with several new developments on the PMT method is given elsewhere [14].

Acknowledgement: We appreciate discussions with Profs. T.Oguchi, S.Bluegel, and P.Blochel. This work was supported by Grant-in-Aid for Scientific Research 23104510.

Appendix A: the error due to the separable form

To evaluate matrix element of a quasilocal operator $X(\mathbf{r}, \mathbf{r}')$, we use separable form $0X0' + 1X1' - 2X2'$ instead of $(0 + 1 - 2)X(0' + 1' - 2')$ under the condition (A') (see Sec.II A). Here 0, 1, 2 means the three components of a eigenfunction as a 3-component function defined in Sec.II A, $0', 1', 2'$ as well.

We have an error because of the separable form. Here we reorganize the discussion to evaluate the error [6, 11] to fit to the formalism in this paper. The error can be evaluated with an identity as;

$$(0 + 1 - 2)X(0' + 1' - 2') - (0X0' + 1X1' - 2X2') = (0 - 2)X(1' - 2') + (1 - 2)X(0' - 2'), \quad (\text{A1})$$

Let us examine the error as the right-hand side of Eq. (A1) under the assumption that X is nearly spherical. Remember that $(0 - 2)$ is completely zero if the condition (A) is satisfied. When the condition (A') is satisfied instead, i.e., when we introduce the finite truncation parameters $l_{\text{max},a}$ and $k_{\text{max},a}$ (given after Eq. (4)), we can expect that $(0 - 2)$ should contain high-energy remnant (high angular momentum l or highly oscillating remnant) with a small amplitude. The remnant $(0 - 2)$ for each L is largest at the MT boundaries. In contrast, when (A') is satisfied, $(1' - 2')$ is low energy part which converges

quickly on the truncation parameters. The value and slope of $(1' - 2')$ are zero at MT boundaries. Thus we can expect the product $(0 - 2)(1' - 2')$ should be small and nearly orthogonal, i.e., $\delta n_a(\mathbf{r}) = (0 - 2)_a(1' - 2')_a$ should satisfy $\int_a d^3r \delta n_a(\mathbf{r}) Y_L(\hat{\mathbf{r}}) \approx 0$ for low L . Here suffix a means quantities within MT at \mathbf{R}_a . Based on these considerations we expect that the error affects little the total energy. This can be checked by changing the truncation parameters within the PMT method.

This logic is applicable not only to the products of the eigenfunctions, but also to the electron density for the Coulomb interaction with some modifications.

Appendix B: Atomic Force

First, we define the Harris energy E_{Harris} [9, 30] which is the total energy of a functional of the density; this gives a reasonable estimate of the total energy even when the density is somehow different from the converged density. When not being converged yet, the input density n^{in} must be treated as one generating the one-particle potential V , and

output density n^{out} which is given from eigenfunctions obtained from the eigenvalue problem of V . Here, V is given by Eq. (34). Now, E_{Harris} in the frozen core approximation as a functional of n^{in} is defined by [9]:

$$E_{\text{Harris}} = E_{\text{k}}^{\text{core}} + E_{\text{B}} - V[n^{\text{Zc}} + n^{\text{in}}, \mathbf{R}_a] \cdot n^{\text{in}} + E_{\text{es}}[n^{\text{Zc}} + n^{\text{in}}, \mathbf{R}_a] + E_{\text{xc}}[n^{\text{Zc}} + n^{\text{in}}], \quad (\text{B1})$$

$$E_{\text{B}} = \sum_p^{\text{occupied}} \alpha_p^{i*} \langle F_i | H^{\text{in}} | F_j \rangle \alpha_p^j, \quad (\text{B2})$$

where E_{B} is the band energy. α_p^i is the eigenvector of $\langle F_i | H^{\text{in}} | F_j \rangle = \langle F_i | \frac{-\Delta}{2m} + V[n^{\text{Zc}} + n^{\text{in}}, \mathbf{R}_a] | F_j \rangle$. Thus we have $E_{\text{B}} = \sum_p^{\text{occupied}} \epsilon_p$, where ϵ_p are eigenvalues. The \mathbf{R}_a -dependence explicitly shown in Eq. (B1) is through the \mathcal{M} -transformation and \mathcal{A} -mapping; even when $n^{\text{Zc}} + n^{\text{in}}$ is fixed as a 3-component function, \mathbf{R}_a -dependence is introduced to Eq. (34) through Eqs.(28,35). In addition, we have \mathbf{R}_a -dependence through $n^{\text{Zc}} + n^{\text{in}}$.

Atomic forces are given as the change of the total energy for atomic displacement $\delta \mathbf{R}_a$. Here let us consider the change of E_{Harris} , written as δE_{Harris} . To obtain δE_{Harris} , we use the derivative chain rule where we treat E_{Harris} as a function of \mathbf{R}_a through $\{F_i(\mathbf{r}), n^{\text{in}}, V, \mathbf{R}_a\}$; V means $V[n^{\text{Zc}} + n^{\text{in}}, \mathbf{R}_a]$ in Eqs.(B1,B2). Remember that there is \mathbf{R}_a dependence through n^{Zc} . Here we assume the partial waves ($\{\phi_{at}(r), \dot{\phi}_{at}(r), \phi_{at}^{\text{Lo}}(r)\}$ in the case of the PMT method) are not dependent on atomic positions as in Ref.[9].

Let us evaluate δE_{Harris} . As for $E_{\text{B}} = \sum_p^{\text{occupied}} \epsilon_p$ as a functional of $\{F_i(\mathbf{r}), V\}$, perturbation theory on Eq. (36) gives

$$\delta E_{\text{B}} = \sum_p^{\text{occupied}} \delta \epsilon_p = \sum_p^{\text{occupied}} \sum_i \sum_j \alpha_{p*}^i (\delta H_{ij} - \epsilon_p \delta O_{ij}) \alpha_p^j = \delta V \cdot n^{\text{out}} + \delta E_{\text{B}}^{\text{Puley}}, \quad (\text{B3})$$

$$\delta E_{\text{B}}^{\text{Puley}} = \sum_p^{\text{occupied}} \sum_i \sum_j \alpha_p^{i*} (\delta H_{ij}^F - \epsilon_p \delta O_{ij}^F) \alpha_p^j, \quad (\text{B4})$$

where we have used $\delta(V \cdot F_i^* F_j) = \delta V \cdot F_i^* F_j + V \cdot \delta(F_i^* F_j)$. $\delta E_{\text{B}}^{\text{Puley}}$ is calculated from $\delta F_{0i}(\mathbf{r})$ and δC_{akL}^i , which are given as a functional of $\delta \mathbf{R}_a$.

Since $E_{\text{es}} + E_{\text{xc}}$ is a functional of $\{n^{\text{in}}, \mathbf{R}_a\}$, we have

$$\begin{aligned} \delta E_{\text{Harris}} &= \delta E_{\text{B}} - \delta(V \cdot n^{\text{in}}) + \delta(E_{\text{es}} + E_{\text{xc}}) \\ &= \delta V \cdot (n^{\text{out}} - n^{\text{in}}) + \delta E_{\text{B}}^{\text{Puley}} + \left. \frac{\partial(E_{\text{es}} + E_{\text{xc}})}{\partial \mathbf{R}_a} \right|_{n^{\text{in}}} \delta \mathbf{R}_a. \end{aligned} \quad (\text{B5})$$

There are three terms in the right hand side of Eq. (B5). The first term appears because E_{Harris} is not converged yet.

To calculate the first term, we need to know δn^{in} which determines δV . When the self-consistency is attained and converged, that is, $n^{\text{in}} = n^{\text{out}}$, $\delta \mathbf{R}_a$ uniquely determines $\delta n^{\text{in}} = \delta n^{\text{out}}$. However, this is not true when $n^{\text{in}} \neq n^{\text{out}}$. In this case, there is no unique way to determine δn^{in} for given $\delta \mathbf{R}_a$. Thus we need an extra assumption to determine it. As a reasonable and convenient choice, we use $\delta n^{\text{in}} = 0$ in the sense of 3-component representation. Physically, this means that $n_{1,a}(\mathbf{r}) - n_{2,a}(\mathbf{r})$ together with frozen core centered at \mathbf{R}_a moves rigidly to $\mathbf{R}_a + \delta \mathbf{R}_a$. Then we can calculate corresponding δV through the change $\delta \bar{n}_0^{\text{Zcv}}$ in Eq. (34). $\delta \bar{n}_0^{\text{Zcv}}$ is evaluated from Eq. (28), where note that $n_0^{\text{Zc}}(\mathbf{r})$ contains \mathbf{R}_a dependence as given in Eq. (23).

Appendix C: onsite matrix

Here we summarize expressions of one-center matrix for O_{ij} , T_{ij} , and V_{ij} . These are essentially the same as what is shown in Ref.[5]. With the help of Eqs.(4,5), Eqs(9,10,34) are reduced to be

$$O_{ij} = \int_{\Omega} d^3r F_{0i}^*(\mathbf{r}) F_{0j}(\mathbf{r}) + \sum_{akk'L} C_{akL}^{*i} \sigma_{akk'L} C_{akL}^j \quad (\text{C1})$$

$$T_{ij} = \frac{1}{2m_e} \int_{\Omega} d^3r \nabla F_{0i}^*(\mathbf{r}) \nabla F_{0j}(\mathbf{r}) + \sum_{akk'L} C_{akL}^{*i} \tau_{akk'L} C_{akL}^j, \quad (\text{C2})$$

$$V_{ij} = \int_{\Omega} d^3r F_{0i}^*(\mathbf{r}) V_0(\mathbf{r}) F_{0j}(\mathbf{r}) + \sum_{akk'LL'} C_{akL}^{*i} \pi_{akk'LL'} C_{ak'L'}^j, \text{ where} \quad (\text{C3})$$

$$\sigma_{akk'l} = \int_{|\mathbf{r}| \leq R_a} d^3r \left(\tilde{P}_{akL}(\mathbf{r}) \tilde{P}_{ak'L}(\mathbf{r}) - P_{akL}(\mathbf{r}) P_{ak'L}(\mathbf{r}) \right), \quad (C4)$$

$$\tau_{akk'l} = \frac{1}{2m_e} \int_{|\mathbf{r}| \leq R_a} d^3r \left(\nabla \tilde{P}_{akL}(\mathbf{r}) \nabla \tilde{P}_{ak'L}(\mathbf{r}) - \nabla P_{akL}(\mathbf{r}) \nabla P_{ak'L}(\mathbf{r}) \right), \quad (C5)$$

$$\pi_{akk'LL'} = \sum_M Q_{kk'LL'M} \mathcal{Q}_{aM}^v + (\tilde{n}_{1,a}^{\text{Zcv}} \circ w + v_{1,a}^{\text{xc}}) \circ \tilde{P}_{akL}(\mathbf{r}') \tilde{P}_{ak'L}(\mathbf{r}') - (\tilde{n}_{2,a}^{\text{Zcv}} \circ w + v_{2,a}^{\text{xc}}) \circ P_{akL}(\mathbf{r}') P_{ak'L}(\mathbf{r}'), \quad (C6)$$

$$Q_{kk'LL'M} = \int_{|\mathbf{r}| \leq R_a} d^3r \left(\tilde{P}_{akL}(\mathbf{r}) \tilde{P}_{ak'Li}(\mathbf{r}) - P_{akL}(\mathbf{r}) P_{ak'Li}(\mathbf{r}) \right) \mathcal{Y}_M(\mathbf{r}). \quad (C7)$$

Note that $\sigma_{akk'l}$ and $\tau_{akk'l}$ are dependent only on l of $L = (l, m)$. In Ref.[5], this $\pi_{akk'LL'}$ is further divided as $\pi_{akk'LL'}^{\text{mesh}} + \pi_{akk'LL'}^{\text{local}}$. \mathcal{Q}_{aM}^v is given by Eq. (35).

Appendix D: scalar relativistic approximation in the augmentation

Roughly speaking, it is allowed to take the scalar relativistic (SR) approximation (e.g. see [2]) if we can safely replace the non-relativistic (NR) wavefunctions with the SR wavefunctions within MTs. The SR wavefunctions contain major and minor components. The major component should be smoothly connected to the NR wavefunction in the interstitial region, where the minority parts are negligible. All physical quantities within MT should be evaluated through the SR wavefunctions. In the followings, we explain how the above idea can be implemented in the 3-component augmentation for bilinear products.

First, we modify the 1st component of the basis. We use two component wavefunctions $\{\mathbf{g}_{1i,aL}(\mathbf{r}), \mathbf{f}_{1i,aL}(\mathbf{r})\}$ instead of $F_{1i,a}(\mathbf{r})$, where the SR approximation gives

$\mathbf{f}_{1i,a}(\mathbf{r}) = \frac{1}{2m_e c} \frac{d\mathbf{g}_{1i,a}(\mathbf{r})}{dr}$, where c is the speed of light. For given F_{0i} and F_{2i} (they are the same as those of the NR case), we ask the the major components $\mathbf{g}_{1i,a}(\mathbf{r})$ to satisfy the boundary conditions as for value and slope at MT boundaries.

In order to calculate the contributions due to the 1st components within the SR approximation instead of the NR approximation, we make a replacement $F_{1i,a}^*(\mathbf{r}) F_{1j,a}(\mathbf{r}') \rightarrow \mathbf{g}_{1i,a}^*(\mathbf{r}) \mathbf{g}_{1j,a}(\mathbf{r}') + \left(\frac{1}{2m_e c}\right)^2 \mathbf{f}_{1i,a}^*(\mathbf{r}) \mathbf{f}_{1j,a}(\mathbf{r}')$. With this replacement, we can evaluate the density n , the matrix O_{ij} and so on. This ends up with the total energy in the SR approximation.

Finally, we see that changes are in the replacement Eqs.(C4-C7), where products $\tilde{P}_{akL}(\mathbf{r}) \tilde{P}_{ak'L}(\mathbf{r})$ (and those with ∇) should be interpreted not only from the products of the majority wavefunctions, but also from those of the minority. This occurs also for the density $n_{1,a}$ included in Eq. (C6).

In such a way we can include the SR effect in the 3-component formalism. In a similar manner, we can include the spin-orbit coupling in the 1st component, which results in the spin off-diagonal contributions [31].

-
- [1] O. K. Andersen, Phys. Rev. B **12**, 3060 (1975).
 - [2] R. Martin, *Electronic structure: basic theory and practical methods* (Cambridge University Press, 2000).
 - [3] D. J. Singh, *Planewaves, Pseudopotential, and the LAPW Method* (Kluwer Academic, Boston, 1994).
 - [4] S. Blügel and G. Bihlmayer, (2006), <http://www2.fz-juelich.de/nic-series/volume31/bluegel.pdf>.
 - [5] M. Methfessel, M. van Schilfhaarde, and R. A. Casali, in *Lecture Notes in Physics*, Vol. 535, edited by H. Dreyse (Springer-Verlag, Berlin, 2000).
 - [6] P. E. Blochl, Phys. Rev. B **50**, 17953 (1994).
 - [7] D. Singh, Phys. Rev. B **43**, 6388 (1991).
 - [8] T. Kotani and M. van Schilfhaarde, Phys. Rev. B **81**, 125117 (2010).
 - [9] M. Methfessel and M. van Schilfhaarde, Phys. Rev. B **48**, 4937 (1993).
 - [10] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
 - [11] J. M. Soler and A. R. Williams, Phys. Rev. B **40**, 1560 (1989).
 - [12] J. M. Soler and A. R. Williams, Phys. Rev. B **42**, 9728 (1990).
 - [13] J. M. Soler and A. R. Williams, Phys. Rev. B **47**, 6784 (1993).
 - [14] T. Kotani and H. Kino, J. Phys. Soc. Jpn. **82**, 124714 [8 Pages] (2013), WOS:000327350700043.
 - [15] T. Kotani, M. van Schilfhaarde, and S. V. Faleev, Physical Review B **76**, 165106 (2007).
 - [16] M. van Schilfhaarde, T. Kotani, and S. V. Faleev, Phys. Rev. B **74**, 245125 (2006).
 - [17] S. V. Faleev, M. van Schilfhaarde, and T. Kotani, Phys. Rev. Lett. **93**, 126406 (2004).
 - [18] M. van Schilfhaarde, T. Kotani, and S. Faleev, Phys. Rev. Lett. **96**, 226402 (2006).
 - [19] T. Kotani, J. Phys. Soc. Jpn. **83**, 094711 [11 Pages] (2014), WOS:000340822100029.
 - [20] M. J. Han, H. Kino, and T. Kotani, Physical Review B **90** (2014), 10.1103/PhysRevB.90.035127.
 - [21] H. Nagara, T. Ishikawa, and T. Kotani, High Pressure Res. **34**, 215 (2014),

WOS:000337605400008.

- [22] J. Klime, M. Kaltak, and G. Kresse, Physical Review B **90** (2014), 10.1103/PhysRevB.90.075125.
- [23] E. Bott, M. Methfessel, W. Krabs, and P. C. Schmidt, J. Math. Phys. **39**, 3393 (1998).
- [24] K. Iwashita, T. Oguchi, and T. Jo, Phys. Rev. B **54**, 1159 (1996).
- [25] M. van Schilfgaarde, Private communication: the head part of the smooth Hankel functions are difficult to be expanded in the series of polynomials accurately.
- [26] M. Weinert, J. Math. Phys. **22** (1981).
- [27] P. Pulay, Mol. Phys. **17** (1969).
- [28] D. J. S. E. Sjøstedt, L. Nordstrom, Solid State Communications **114**, 15 (2000).
- [29] C. Filippi, D. J. Singh, and C. J. Umrigar, Phys. Rev. B **50**, 14947 (1994).
- [30] J. Harris, Phys. Rev. B **31**, 1770 (1985).
- [31] A. N. Chantis, M. van Schilfgaarde, and T. Kotani, Phys. Rev. Lett. **96**, 086405 (2006).